Chapter 7

Inferences Based on a Single Sample: Estimation with Confidence Intervals

Identifying the Target Parameter

Target parameter – the unknown population parameter of interest for estimating

Parameter	Key Words or Phrases	Type of Data
μ	Mean; average	Quantitative
p	Proportion; percentage; fraction; rate	Qualitative

2

Large-Sample Confidence Interval for a Population Mean

How to estimate the population mean and assess the estimate's reliability?

$\bar{x}$ is an estimate of μ, and we use CLT to assess how accurate that estimate is

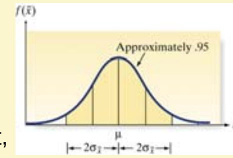
According to CLT, 95% of all $\bar{x}$ from sample size n lie within $\pm 1.96\sigma_{\bar{x}}$ of the mean

We can use this to assess accuracy of $\bar{x}$ as an estimate of μ

3

Large-Sample Confidence Interval for a Population Mean

$$\bar{x} \pm 2\sigma_{\bar{x}} = \bar{x} \pm \frac{2\sigma}{\sqrt{n}} \approx .95$$



We are about 95% confident, for any $\bar{x}$ from sample size n, that μ will lie in the interval $\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$

4

Large-Sample Confidence Interval for a Population Mean

We usually don't know σ, but with a large sample *s* is a good estimator of σ.

We can calculate confidence intervals for different confidence coefficients.

Interval Estimator (or confidence interval) – a formula used to calculate an interval estimate from sample data

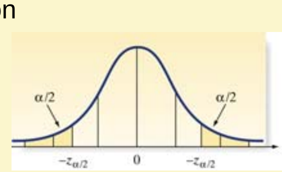
Confidence coefficient – probability that a randomly selected confidence interval encloses the population parameter

Confidence level – Confidence coefficient expressed as a percentage

5

Large-Sample Confidence Interval for a Population Mean

The confidence coefficient is equal to 1-α , and is split between the two tails of the distribution



6

Large-Sample Confidence Interval for a Population Mean

The Confidence Interval is expressed more generally as

$$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}} = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

For samples of size > 30, the confidence interval is expressed as

$$\bar{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Requires that the sample used be random

7

Large-Sample Confidence Interval for a Population Mean

TABLE 7.2 Commonly Used Values of $z_{\alpha/2}$

Confidence Level $100(1-\alpha)$	α	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.96
99%	.01	.005	2.575

8

Small-Sample Confidence Interval for a Population Mean

2 problems presented by sample sizes of less than 30

- CLT no longer applies
- Population standard deviation is almost always unknown, and s may provide a poor estimation when n is small

9

Small-Sample Confidence Interval for a Population Mean

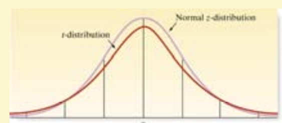
If we can assume that the sampled population is approximately normal, then the sampling distribution of $\bar{x}$ can be assumed to be approximately normal

Instead of using $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ we use $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

This *t* is referred to as the t-statistic

10

Small-Sample Confidence Interval for a Population Mean



The t-statistic has: a sampling distribution very similar to *z*

Variability dependent on n, or sample size.

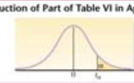
Variability is expressed as (n-1) degrees of freedom (df). As (df) gets smaller, variability increases

11

Small-Sample Confidence Interval for a Population Mean

- Table for t-distribution contains t-value for various combinations of degrees of freedom and t_α
- Partial table here shows components of table

TABLE 7.3 Reproduction of Part of Table VI in Appendix A



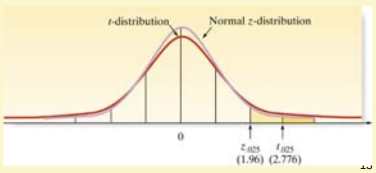
Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

12

Small-Sample Confidence Interval for a Population Mean

Comparing t and z distributions for the same α, with df=4 for the t-distribution, you can see that the t-score is larger, and therefore the confidence interval will be wider.

The closer df gets to 30, the more closely the t-distribution approximates the normal distribution



Small-Sample Confidence Interval for a Population Mean

When creating a Confidence interval around μ for a small sample we use

$$\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

basing $t_{\alpha/2}$ on n-1 degrees of freedom

We assume a random sample drawn from a population that is approximately normally distributed

14

Large-Scale Confidence Interval for a Population Proportion

Confidence intervals around a proportion are confidence intervals around the probability of success in a binomial experiment

Sample statistic of interest is $\hat{p}$

Mean of sampling distribution of $\hat{p}$ is p. p is an unbiased estimator of $\hat{p}$

Standard deviation of the sampling distribution is $\sigma_{\hat{p}} = \sqrt{pq/n}$ where q=1-p

For large samples, the sampling distribution of $\hat{p}$ is approximately normal

Large-Scale Confidence Interval for a Population Proportion

Sample size n is large if $\hat{p} \pm 3\sigma_{\hat{p}}$ falls between 0 and 1

Confidence interval is calculated as

$$\hat{p} \pm z_{\alpha/2}\sigma_p = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{pq}{n}} \approx \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where $\hat{p} = \frac{x}{n}$ and $\hat{q} = 1 - \hat{p}$

16

Large-Scale Confidence Interval for a Population Proportion

When p is near 0 or 1, the confidence intervals calculated using the formulas presented are misleading. An adjustment can be used that works for any p, even with very small sample sizes

Agresti and Coull Correction:

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

17

Agresti and Coull Correction:

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}};$$

Where, $\tilde{p} = \frac{x+2}{n+4}$;

x = number of successes;

n = sample size

18